

An Integrated, Distributed Software System to Support DNA Sequencing and Analysis

Arthur Kobayashi, David Ow, Tom Slezak, Mark Wagner, Mimi Yeh, Jane Lamerdin, Paula M. McCreedy, Anthony V. Carrano

We have been developing an integrated, distributed software system to support our sequencing and analysis efforts at LLNL. The system consists of a diverse suite of software programs, developed both in-house and elsewhere, that run on an extensive network of Sun workstations and Macintosh computers. The system components are integrated through our Sybase relational database. The software has been implemented using a variety of programming languages and makes use of many external packages and tools. The system performs a number of functions to facilitate our sequencing process and manage sequencing data:

- track sequencing libraries and DNA samples.
- streamline DNA sequencer setups.
- track sequencer data output and move sequence data to assembly directories.
- assemble and display sequence contigs, specify sequences and protocols for closure.
- submit and track sequence feature searches, process search results.
- archive and display features for analysis and annotation, streamline submissions.

In addition, we have developed tools to track and display sequencing library status (where each library is in the sequencing process) and to coordinate work between various groups. Data is transferred between processes using scripts, and information is archived in the database at strategic points in the overall process. In the database, analysis results are linked with our existing repository of physical map data.

We are developing software using a variety of languages and toolkits, including: C, X Windows/Motif, Perl, Hypertext Markup Language (HTML), Unix scripts, customized spreadsheets, and UIM/X (a commercial graphical user-interface builder). We have also evaluated a number of commercial and community-provided programs which have been incorporated into our system.

This system greatly facilitates the flow of information and materials through our facilities, and provides a very comprehensive view of information linking physical map and sequencing analysis data.

This work was performed by Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy, Contract No. W-7405-Eng-48.